Title
Detecting Coronavirus Disease 2019 Pneumonia in Chest X-Ray Images Using Deep Learning

Author
Zhu, Ziqi

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Detecting Coronavirus Disease 2019 Pneumonia in Chest X-Ray Images Using Deep Learning

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Ziqi Zhu

2020

ABSTRACT OF THE THESIS

Detecting Coronavirus Disease 2019 Pneumonia
in Chest X-Ray Images
Using Deep Learning

by

Ziqi Zhu
Master of Science in Statistics
University of California, Los Angeles, 2020
Professor Yingnian Wu, Chair

The coronavirus disease 2019 (COVID-19) pandemic has already become a global threat. To fight against COVID-19, effective and fast screening methods are needed. This study focuses on leveraging deep learning techniques to automatically detect COVID-19 pneumonia in chest X-ray images. Two models are trained based on transfer learning and residual neural network. The first one is a binary classifier that separates COVID-19 pneumonia and non-COVID-19 cases. It classifies all test cases correctly. The second one is a four-class classifier that distinguishes COVID-19 pneumonia, viral pneumonia, bacterial pneumonia and normal cases. It reaches an average accuracy, precision, sensitivity, specificity, and F1-score of 93%, 93%, 93%, 97%, and 93%, respectively. To understand on how the four-class classifier detects COVID-19 pneumonia, we apply Gradient-weighted Class Activation Mapping (Grad-CAM) method and find out that the classifier is able to focus on the patchy areas in chest X-ray images and make accurate predictions.

The thesis of Ziqi Zhu is approved.

Chad J. Hazlett

Hongquan Xu

Yingnian Wu, Committee Chair

University of California, Los Angeles

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my Mom and Dad, for their unfailing support and unwavering love.

CHAPTER 1

Introduction

The novel coronavirus disease 2019 (COVID-19) pandemic is an ongoing pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of 28 May 2020, more than 5.5 million cases of COVID-19 have been reported in 216 countries, areas and territories, resulting in 353,373 deaths [Org20]. While most people only have mild to moderate symptoms, some patients have developed severe illnesses that include pneumonia and acute respiratory distress syndrome (ARDS). To control the spreading of COVID-19, effective screening of patients is critical. So far, the gold standard screening method is the reverse transcription polymerase chain reaction (RT-PCR) test which has been designed to detect SARS-CoV-2 genetically. But it only has a positive rate ranging between 30% and 60% [AYH20, YYS]. For patients who develop severe illnesses such as pneumonia and ARDS, a good complementary screening method is radiography examination, where chest radiography imaging (e.g., X-ray or computed tomography (CT) scan) is analyzed for SARS-CoV-2 viral infection indicators including bilateral, peripheral ground-glass opacities and pulmonary consolidations [SAB20]. However, the viral infection indicators can be subtle and it is difficult for radiologists to distinguish COVID-19 pneumonia from normal cases or other pneumonias. Thus, computer-aided diagnostic systems that can help detect COVID-19 pneumonia in chest radiography images are highly desired.

In deep learning area, convolutional neural networks (CNN) are typically used to solve object detection and image classification problem due to its ability to preserve spatial structure and recognize image features. LeNet-5 [LBB98], introduced by LeCun et al. in 1998, is the first instantiation of CNN that has been successfully used in practice. With only seven

layers, it has achieved high performance for digit recognition problem. In 2012, AlexNet [KSH12] entered the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [RDS15], a benchmark for object detection and image classification tasks at large scale, and was able to outperform all previous non-deep-learning-based models by a significant margin with top-5 error rate at 16.4%. In 2013, ZFNet [ZF14] was designed with the architecture of AlexNet and better hyperparameters. Thus, it reached less top-5 error rate at 11.7%. In 2014, VGGNet [SZ14] and GoogLeNet [SLJ15] both made another jump in performance, with top-5 error rate at 7.3% and 6.7%, respectively. The common thing between VGGNet and GoogLeNet is that they both have deeper architectures. In 2015, the ILSVRC winner is the residual neural network (ResNet) with 152 layers [HZR16]. It achieved the top-5 error at 3.6% which is less than human error 5.1%. This is the first time that a CNN model outperforms human in large scale object recognition tasks and it marks the promising future for CNN models.

Motivated by the need for fast and accurate analysis of radiography images, a number of COVID-19 pneumonia detection models based on state-of-the-art CNN models have been proposed and results have shown to be promising [CRK20, NKP20, AM20]. However, those studies are reporting the results on small datasets under binary classification (COVID-19 pneumonia and non-COVID-19 cases) or three-class classification (COVID-19 pneumonia, other pneumonias and normal cases) problem. In this study, we have prepared a comparatively large dataset consisting chest X-rays images of COVID-19 pneumonia, viral pneumonia, bacterial pneumonia, and normal cases, and have trained two models based on transfer learning concept and ResNet architecture. One model is trained to distinguish COVID-19 pneumonia and non-COVID-19 cases in chest X-ray images, while the other one is trained to classify COVID-19 pneumonia, viral pneumonia, bacterial pneumonia, and normal cases. Using the first model, we can quickly screen patients who are suspected of having COVID-19 pneumonia and arrange immediate medical cares for them. The second model can help clinicians to not only better screen COVID-19, but also decide treatment strategy and make treatment plan based on cause of the infection.

This thesis is organized as follows. First, Chapter 2 describes the generation of dataset

for two models. Chapter 3 describes the model architecture, transfer learning concept, implementation details, performance evaluation metrics, and the model interpretation strategy. Chapter 4 presents and discusses the results of experiments conducted to evaluate and interpret the proposed models. Finally, conclusions are drawn in Chapter 5.

CHAPTER 2

Dataset

2.1 Data Description

The dataset used to train and evaluate our models consists 6,216 chest X-ray images across 6,001 patients. To generate this dataset, we combine and modify two different datasets that are publicly available: COVID Chest X-Ray Dataset [CMD20] and Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images [KZG18]. The following section gives a description of the two datasets and a summary of how we generate our dataset:

COVID Chest X-Ray Dataset: This dataset comprises 360 chest X-ray images and CT scans across 198 patients which are positive or suspected of COVID-19 pneumonia or other pneumonias. In this study, we only select 232 chest X-ray images with anteroposterior or posteroanterior chest view from 145 patients who are positive or suspected of COVID-19 pneumonia.

Labeled OCT and Chest X-Ray Images: This dataset has a total of 109,309 OCT images and 5,856 chest X-ray images that have been collected and labeled by the same lab. The chest X-ray images include 4,273 pneumonia images (1,493 viral and 2,780 bacterial) and 1,583 normal images from total of 5,856 patients. X-ray images are already split into training set and testing set. In this study, we choose all chest X-ray images and use the predefined training set and testing set.

The summary of our dataset is shown in Table 2.1. Sample chest X-ray images for each pneumonia type are shown in Figure 2.1.

Table 2.1: Details of our dataset

Original Dataset	Type	Total Patients	Total Images
COVID Chest X-Ray Dataset	COVID-19	145	232
Labeled OCT and Chest X-Ray Images	Viral	1493	1493
	Bacterial	2780	2780
	Normal	1583	1583

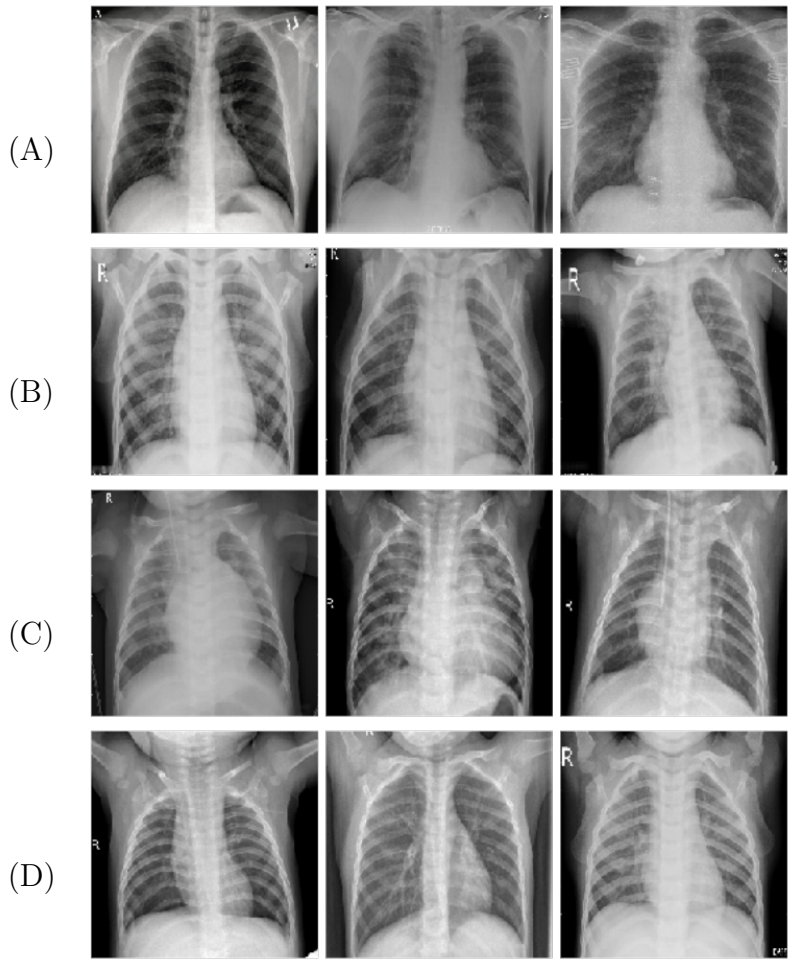

Figure 2.1: Sample chest X-ray images from our dataset: (A) COVID-19 pneumonia, (B) viral pneumonia, (C) bacterial pneumonia, and (D) normal case.

2.2 Dataset Preprocessing

Our dataset consists of 232 COVID-19 pneumonia images, 1,493 viral pneumonia images, 2,780 bacterial pneumonia images and 1,583 normal case images, shown in Table 2.1. For COVID-19 pneumonia images, there are multiple images from some of the patients. In order to ensure that images of each patient appear only in training set or testing set, we separate the data by patient before splitting into training and testing set. Among 232 COVID-19 pneumonia images, 172 images fall into the training set and the rest 60 images fall into the testing set. For chest X-ray images of the other pneumonia types, we use the predefined training set and testing set in the "Labeled OCT and Chest X-Ray Images" dataset in our four-class classification model, and label all of them as non-COVID-19 cases in our binary classification model. The data distributions of the binary classification problem and four-class classification problem are shown in Figure 2.2.

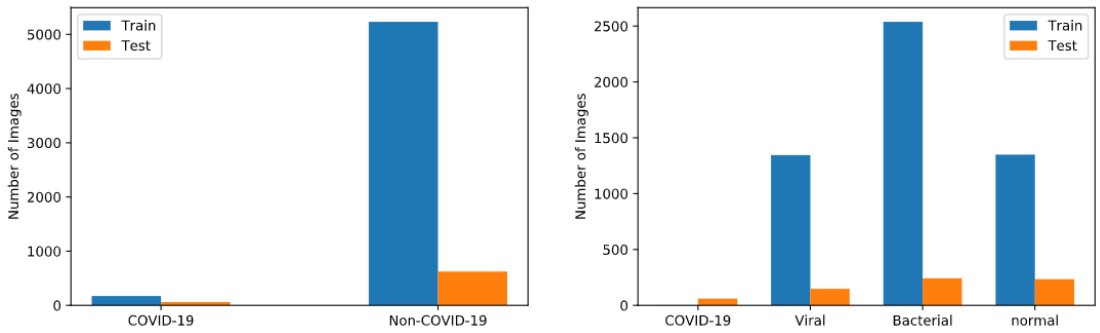

Figure 2.2: Data distribution for: (A) binary classification, and (B) four-class classification.

A noticeable problem is the limited amount of COVID-19 pneumonia images, which causes the data to be imbalanced and adds challenge to the performance of our classification models. To alleviate this problem, we use over-sampling method to even-up the data between classes. More specifically, we make COVID-19 pneumonia images 8-fold in the four-class classification problem and 24-fold in the binary classification problem. The details of our final training sets and testing sets are shown in Table 2.2 and Table 2.3. The final data distribution of two

problems are shown in Figure 2.3.

Table 2.2: Details of training and testing set for the binary classification model

Label	Pneumonia Type	Training Set	Testing Set	Total
Non-COVID-19	Viral, Bacterial, and Normal (no infection)	5232	624	5856
COVID-19	COVID-19	4128	60	4188

Table 2.3: Details of training and testing set for the four-class classification model

Label	Pneumonia Type	Training Set	Testing Set	Total
COVID-19	COVID-19	1376	60	1436
Viral	Viral	1345	148	1493
Bacterial	Bacterial	2538	242	2780
Normal	Normal (no infection)	1349	234	1583

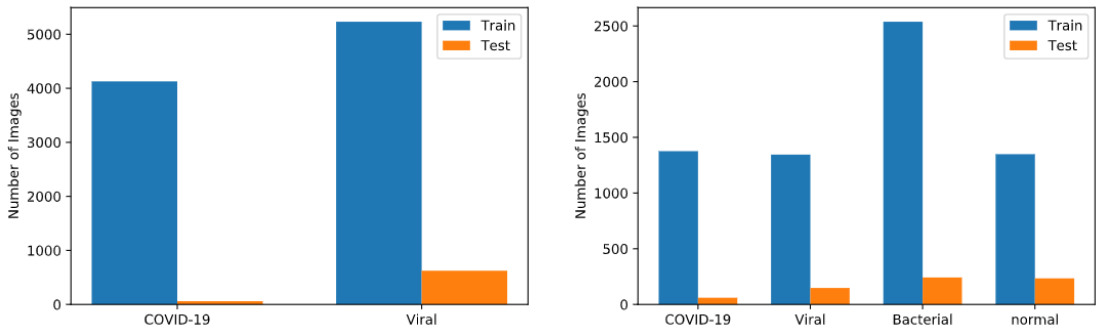

Figure 2.3: Final data distribution for: (A) binary classification, and (B) four-class classification.

CHAPTER 3

Methodology

In this chapter, we will discuss the model architecture, transfer learning concept, implementation details, performance evaluation metrics, and the strategy for interpreting models.

3.1 ResNet Architecture

Residual neural network (ResNet) is proposed by He et al. in 2015 [HZR16]. The hypothesis behind ResNet is that deeper networks are harder to optimize, since the deeper model should be able to perform as well as the shallower model by copying the learned parameters from the shallower model and setting additional layers to identity mapping. To help optimize deeper models, residual blocks are designed to fit a residual mapping $F(x)$ instead of the desired underlying mapping $H(x)$, and full ResNet architecture is built by stacking residual blocks. More specifically, every residual block has two 3×3 convolutional layers. Periodically, the number of filters are doubled and spatial downsampling is operated. Figure 3.1 illustrates the structure within the residual block and Figure 3.2 shows an example of ResNet architecture - ResNet with 18 layers (ResNet18).

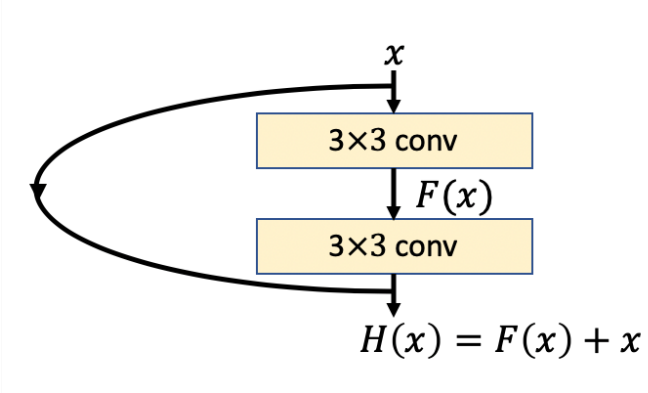

Figure 3.1: Structure of residual block

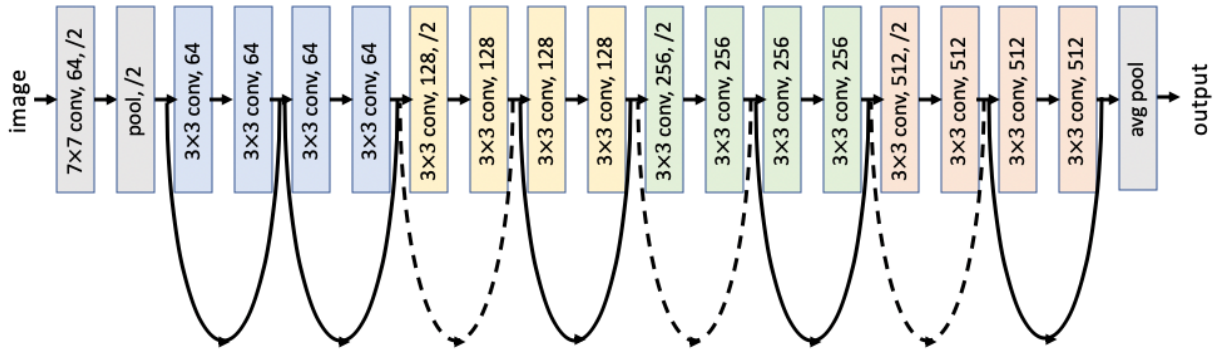

Figure 3.2: Schema of ResNet18 architecture

In this study, we build two deep models based on ResNet18 for the classification under two schemes. One is a binary classification that distinguishes COVID-19 pneumonia and non-COVID-19 cases, while the other is a four-class classification that classifies COVID-19 pneumonia, viral pneumonia, bacterial pneumonia, and normal cases.

3.2 Transfer Learning

A successful training of deep neural networks often requires a large-scale dataset and a long training period. Moreover, a basic assumption for many deep learning models is that the

training and testing data should be drawn from the same distribution. In many real-word applications, the availability of data is limited due to the high cost of collecting and labelling data. In such cases, retaining and reusing previously learned knowledge for a different data distribution, task or domain is critical. Transfer learning is a machine learning method where a previously trained model is reused as a starting point for the new model and task. It can not only alleviate the need and effort to collect training data, but also accelerate training process.

There are two commonly used strategies to exploit transfer learning on deep convolutional neural network. The first strategy is called feature extraction where the pretrained model, retaining both its initial architecture and the learned parameters, is only used to extract image features for the input of the new classification model. The second strategy makes some modifications to the pretrained model, such as architecture adjustments and parameter tuning, to improve extracted image features on the new dataset and achieve optimal results.

In this study, the second strategy is used. More specifically, we use ImageNet[RDS15], a large-scale dataset consisting 1.2 million high-resolution images in 1,000 different classes (i.e.,fish, bird, tree, flowers, sport, room, etc.), to pretrain ResNet18. Since images in ImageNet are different from our dataset, we cannot directly use the pretrained model to extract image features for classification. Instead, we need to fine-tune parameters of the pretrained model on our dataset to generate better image features. Besides, the architecture of ResNet18 has been changed - the number of outputs in the final fully-connected layer is changed from 1,000 to the number of classes in each classification problem. Thus, the fine-tuned model outputs scores for each class and can classify input image to the class with maximum score. The schematic representation of the training process is depicted in Figure 3.3.

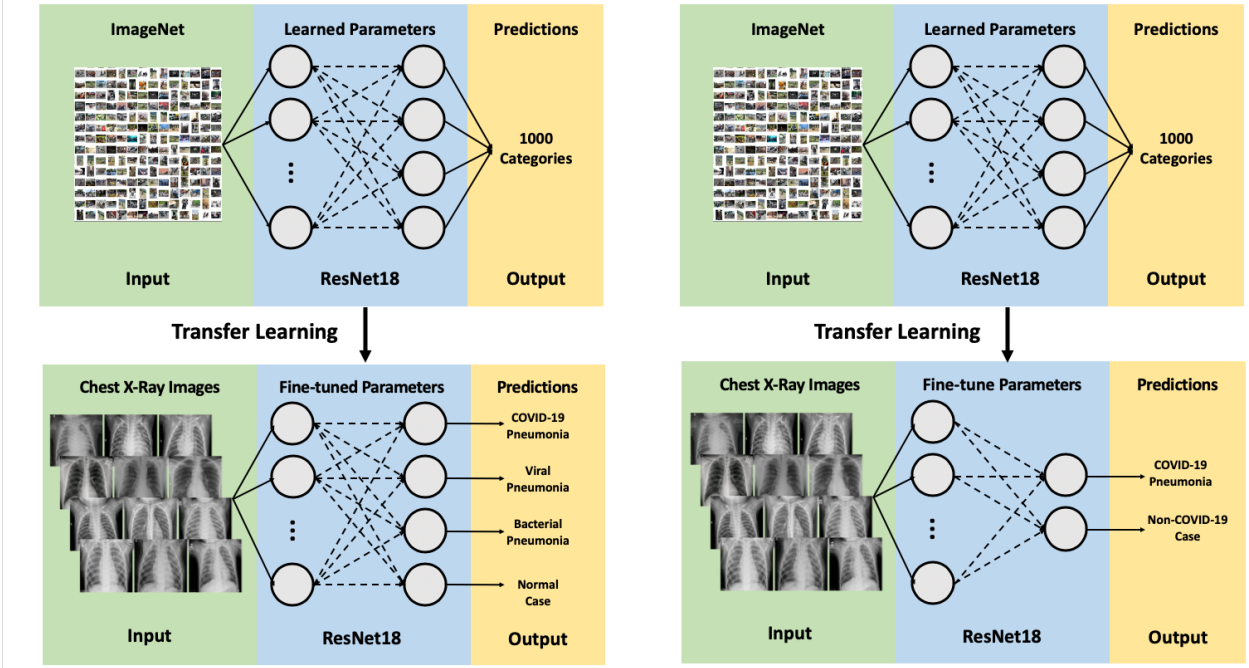

Figure 3.3: Schematic of transfer learning with ResNet18 for: (A) four-class classification problem, and (B) binary classification problem.

3.3 Implementation Details

3.3.1 Data Augmentation

We apply data augmentation methods that include rotation, scaling, translation, and horizontal flip on the training set to address the data deficiency problem. That means, each image in the training set is randomly operated by the following methods:

- **Rotation:** Rotating the image with an angle between 0 and 15 in the clockwise or counter clockwise direction.
- **Scaling:** Sampling the scale of frame size of the image randomly between 90% and 110%.
- **Translation:** Translating image horizontally and vertically between -10% and 10%.
- **Horizontal Flip:** Horizontally flipping the image with a probability of 0.5.

3.3.2 Hyperparameters

In this study, all images are normalized and resized to 224×224 pixels. We use five-fold cross validation to evaluate the model and tune the hyperparameters, and obtain the generalized results in the testing stage. The following hyperparameters are used for training:

Binary classification model: Batch size = 256, cross entropy loss, stochastic gradient descent (SGD) optimizer with momentum = 0.9 and weight decay (L2 penalty) = 0.01, learning rate = 0.003 (will decay by 0.1 at 10 and 20 epoch), 50 epochs in total, and early stopping with patience = 10.

Four-class classification model: Batch size = 256, cross entropy loss, SGD optimizer with momentum = 0.9 and weight decay (L2 penalty) = 0.1, cross entropy loss, learning rate = 0.003 (will decay by 0.1 at 20 and 40 epoch), 50 epochs in total, and early stopping with patience = 15.

3.4 Performance Evaluation Metrics

Due to the data imbalance problem, the performance of the model is evaluated using eight evaluation metrics - accuracy, confusion matrix, precision, recall (or sensitivity), specificity (or selectivity), F1-score, receiver operating characteristic (ROC) curve and area under the curve (AUC).

3.4.1 Confusion Matrix

A confusion matrix is a summary of predicted labels and true labels on a classification problem. It gives us insights not only into the errors but also into the types of errors that are being made by the classifier. The schema of confusion matrix in a binary classification problem is shown in Figure 3.4.

		Predicted Label	
		Positive	Negative
True Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 3.4: Schema of confusion matrix

The accuracy, precision, recall (or sensitivity), specificity (or selectivity) and F1-score can be computed as follows based on the confusion matrix.

$$\text{accuracy} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall (or sensitivity)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{specificity (or selectivity)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

For multi-class classification problem, we can not only use one-vs-all methodology to compute the above metrics but also use macro average and weighted average of metrics to evaluate the overall model performance. The macro average score is average of the target metric (e.g., accuracy) over all classes while the weighted average is average of the target metric over classes weighted by sample size. If each class has the same sample size, the macro average and the weighted average are always the same.

3.4.2 ROC Curve and AUC

ROC curve is a graphical plot that shows the performance of a binary classification model. It is created by plotting true positive rate (TPR) against false positive rate (FPR) at various discrimination thresholds, where TPR is also known as sensitivity or recall, and FPR can be calculated as $(1 - \text{specificity})$. By changing the model's discrimination threshold, confusion matrix may be different and a point in the ROC curve can be plotted.

The ROC curve for a random binary classification model would be a diagonal line from (0,0) to (1,1). Any curve above the diagonal line represents good classification model which is better than random, while curve below the diagonal line means the model is worse than random. A perfect binary classification model would yield a straight line from (0,1) to (1,1), meaning 100% sensitivity and 100% specificity.

AUC refers to the area under the ROC curve, which is always between 0 and 1. Based on the concept of ROC curve, it can be concluded that higher the AUC, better the binary classification model. A random classifier has AUC of 0.5, and a great model has AUC close to 1. With AUC less than 0.5, the model tends to reciprocate the classes. A schema of ROC curve and AUC is shown in Figure 3.5.

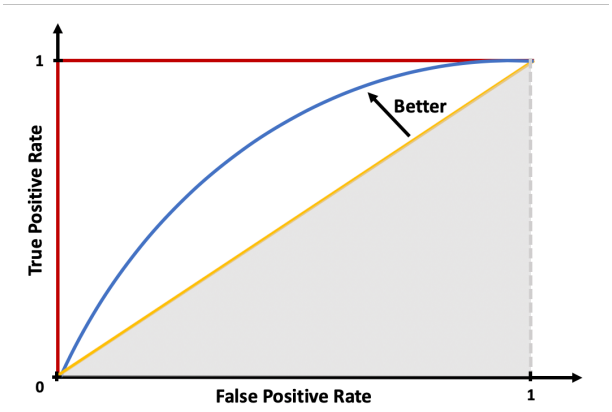

Figure 3.5: Schema of ROC curve and AUC : red line: a perfect classifier, blue curve: a great classifier, yellow line: a random classifier, and shaded area: AUC for the random classifier.

In multi-class classification model, we can plot ROC curve for each class using one-vs-all

methodology.

3.5 Model Interpretation

Gradient-weighted class activation mapping (Grad-CAM) [SCD17] is a deep learning technique that can generate visual explanations of CNN-based models. Thus, it can make models more transparent and easier to interpret. Using the gradients of a target concept (i.e., the COVID-19 class in our four-class classification) flowing into the final convolutional layer, it can produce a localization map showing important regions for the model to make predictions. Grad-CAM not only provides a way to evaluate models, but also helps human to study and understand more about deep learning models.

In our study, we apply Grad-CAM method on the four-class classification model and choose the COVID-19 pneumonia as the target concept to understand how the four-class classification model makes predictions of COVID-19 cases.

CHAPTER 4

Results

In this section, we will discuss performance of the two classification models as well as the interpretation of the four-class classification model.

4.1 Performance of the Binary Classification Model

A binary classifier is implemented to separate COVID-19 pneumonia and non-COVID-19 cases. The classifier manages to classify all COVID-19 and non-COVID-19 test cases correctly. Accuracy, precision, sensitivity, specificity, F1-score and AUC are all 100.0%. Figure 4.1 shows the visualization of the confusion matrix for the binary classification. The summary of results is in Table 4.1 and the ROC curve is shown in Figure 4.2.

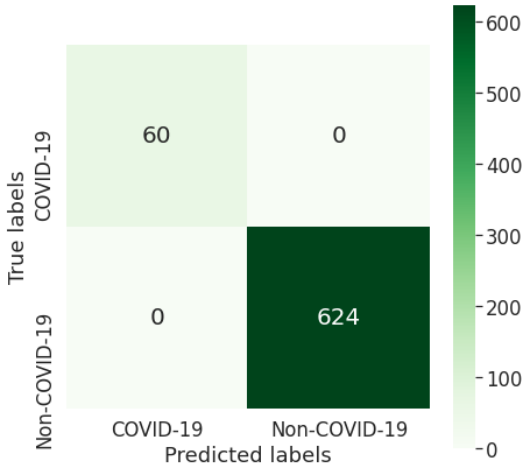

Figure 4.1: Confusion matrix of the binary classification

Table 4.1: Results for binary classification

Label	Size	Accuracy	Precision	Recall	Selectivity	F1-score	AUC
COVID-19	60	1.00	1.00	1.00	1.00	1.00	1.00
Non-COVID-19	624	1.00	1.00	1.00	1.00	1.00	1.00

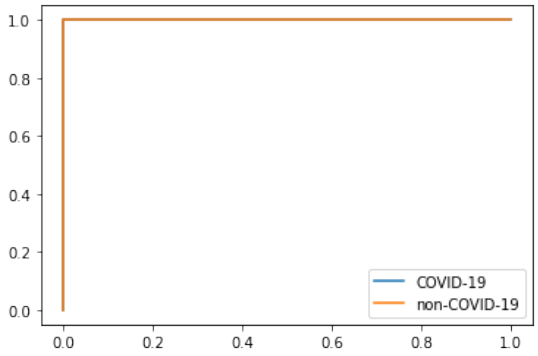

Figure 4.2: ROC curve of the binary classification model

4.2 Performance of the Four-class Classification Model

A four-class classifier is implemented to distinguish COVID-19 pneumonia, viral pneunomia, bacterial pneunomia, and normal cases. This classifier achieves an average accuracy of 93%, with an average precision, sensitivity, specificity, and F1-score of 93%, 93%, 97%, and 93%, respectively. ROC curves are generated to evaluate the model's ability to distinguish each class from other classes using one-vs-all method, and all AUCs are more than 96%. Figure 4.3 shows the visualization of the confusion matrix for the four-class classification model. Table 4.2 summarizes the details of the results for this four-class classification model and Figure 4.4 shows four ROC curves and the zoomed area.

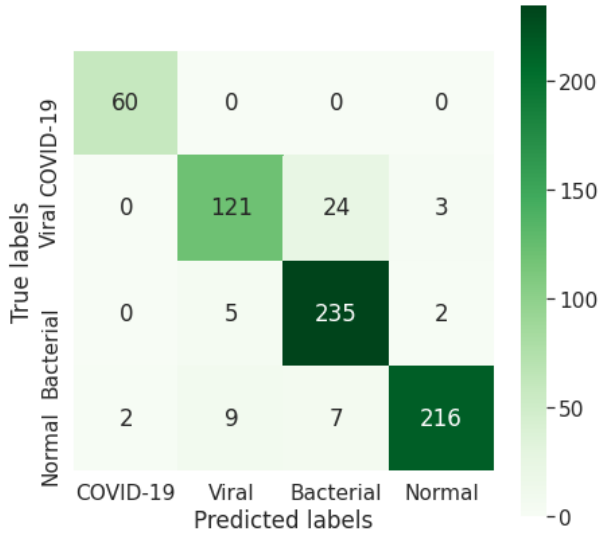

Figure 4.3: Confusion matrix of the four-class classification.

Table 4.2: Details of results for the four-class classification model

Label	Size	Accuracy	Precision	Recall	Selectivity	F1-score	AUC
COVID-19	60	1.00	0.97	1.00	1.00	0.98	1.00
Viral	148	0.82	0.90	0.82	0.97	0.86	0.96
Bacterial	242	0.97	0.88	0.97	0.93	0.93	0.98
Normal	234	0.92	0.98	0.92	0.99	0.95	0.98
Macro avg		0.93	0.93	0.93	0.97	0.93	0.99
Weighted avg		0.92	0.93	0.92	0.97	0.92	0.98

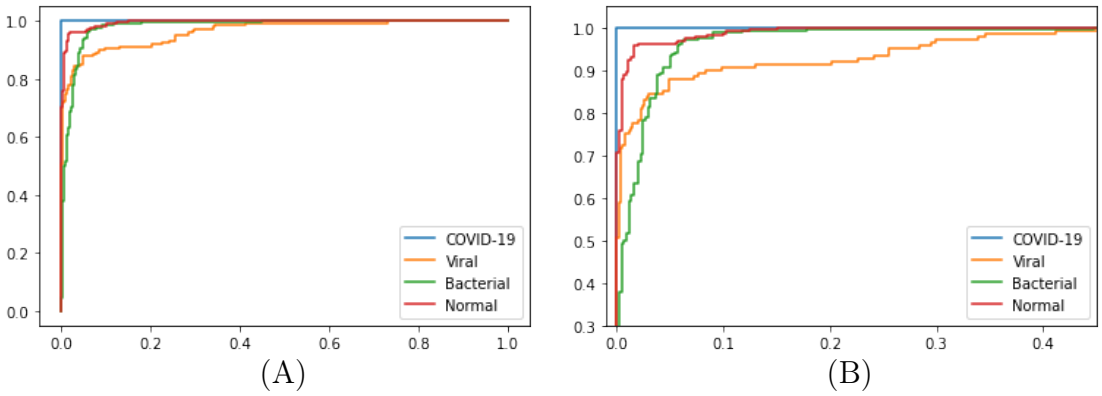

Figure 4.4: ROC curve of the four-class classification model: (A) the original ROC curves, and (B) a zoomed area of the ROC curves.

Overall, the model reaches very high performance in distinguishing four classes. For COVID-19 pneumonia, it is clear that all COVID-19 test cases are classified correctly. This is very important in screening COVID-19 pneumonia. However, the model seems to be less powerful in distinguishing other three categories - it tends to classify normal cases into other categories and classify viral pneumonia images into bacterial pneumonia.

4.3 Model Interpretation

Grad-CAM method can generate a localization map highlighting important regions in image for model to make predictions. After applying Grad-CAM method on COVID-19 images in the four-class classification model, we can try to understand how the model predicts COVID-19 pneumonia given chest X-ray images. We are able to obtain some chest X-ray images of COVID-19 cases with detailed description and diagnosis by radiologists[Rad]. By comparing the diagnosis and the localization map, we can not only get an better understanding of how the model make predictions but also check if the model focuses on the same region as radiologists do. Some samples are shown in Figure 4.5, 4.6, 4.7, 4.8, and 4.9. Two things need to be noted: 1) the left and right side are reversed in the chest X-ray images, and 2) in the following figures, the left ones are original chest X-ray images while the right ones

(resized to 224×224 pixels) show localization maps generated by Grad-CAM. Radiologists' comments are listed in captions.

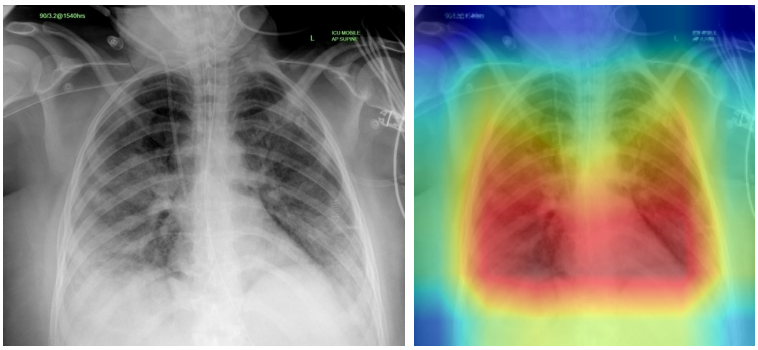

Figure 4.5: Comment: patchy areas of air space opacification bilaterally with a lower zone predominance.

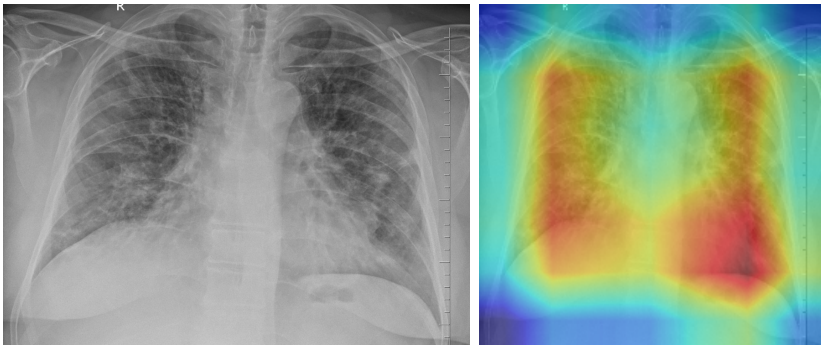

Figure 4.6: Comment: there is peripheral patchy air space opacification seen in both lung lower zones with diffuse ground-glass haze bilaterally.

Figure 4.7: Comment: mutifocal consolidation in right mid zone and left mid/lower zones.

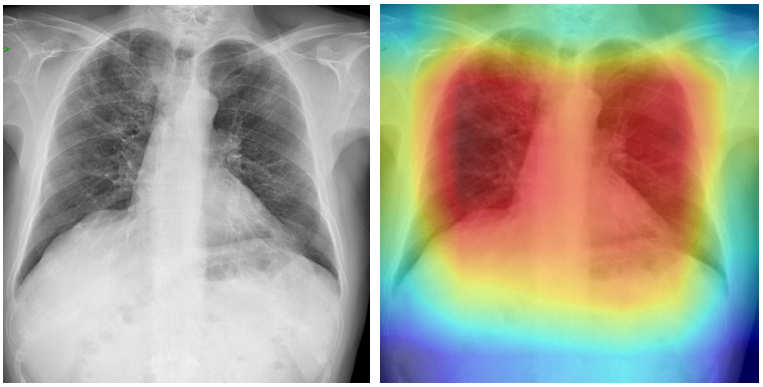

Figure 4.8: Comment: multiple faint alveolar opacities are identified, predominantly peripheral with greater involvement of the upper lobes.

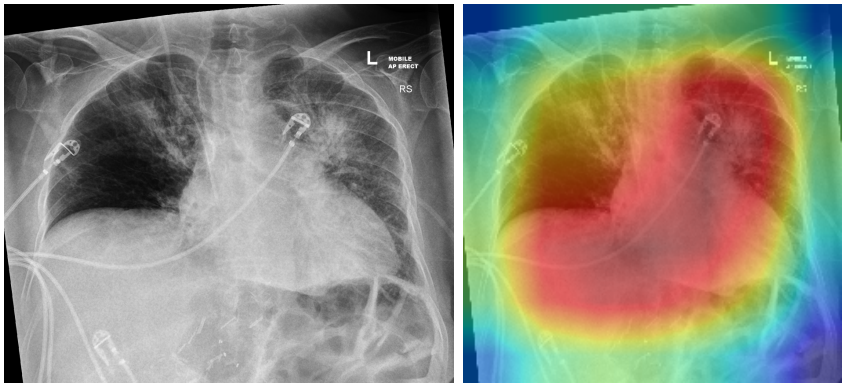

Figure 4.9: Comment: there is established left upper lobe and now progressive patchy consolidation in left lower, right upper and middle lobes.

In Figure 4.5 and 4.6, our model focuses on the right area. In Figure 4.7-4.9, our model tends to be unfocused and highlights larger areas compared to radiologists' descriptions.

From the above figures, we can conclude that 1) our model manages to focus on the lung area instead of other areas (i.e., cervical spine, upper limb, costae, and abdomen) shown in the chest X-ray images, and 2) our model tends to focus more on areas with rich textures (i.e., patchy areas). So, our model is more powerful when the patchy areas are the main indicator of COVID-19 pneumonia. Otherwise, though our model can make accurate predictions, it tends to be unfocused and fails to highlight specific areas that indicating COVID-19 pneumonia.

CHAPTER 5

Conclusion

In this study, we train two deep learning models based on transfer learning concept and ResNet18 architecture for COVID-19 pneumonia detection in chest X-ray images. The first one is a binary classification model that aims to separate COVID-19 pneumonia and non-COVID-19 cases. It is able to classify all test cases correctly. The second one is a four-class classification model that aims to distinguish COVID-19 pneumonia, viral pneumonia, bacterial pneumonia and normal cases. It reaches an average accuracy, precision, sensitivity, specificity, F1-score, and AUC of 93%, 93%, 93%, 97%, 93%, and 99%, respectively. This model is able to detect all COVID-19 test cases, but it tends to classify normal cases into other classes and tends to classify viral pneumonia images into bacterial pneumonia. To shed a light on how the four-class classification model make predictions of COVID-19 pneumonia cases, we apply Grad-CAM method to generate localized map that highlights important regions the model thinks to make predictions. By comparing the highlighted regions and radiologists' descriptions of chest X-ray images, we find out that our model is more powerful when patchy areas are the main indicators of COVID-19 pneumonia. Otherwise, though our model can make accurate predictions, it tends to be unfocused and fails to highlight specific areas that indicating COVID-19 pneumonia.

COVID-19 has already become a global threat and has taken away hundreds of thousands of people's lives. This study shows the feasibility of building a computer-aided diagnostic system that can help clinicians detect COVID-19 pneumonia from radiology images accurately and quickly. Moreover, model interpretation techniques allow us to further evaluate and understand models. However, the limited data adds challenge to the performance of model.

By collecting more chest X-ray images of COVID-19 pneumonia, other pneumonias and normal cases, the model will be more robust and powerful.

REFERENCES

[AKM17] Benjamin Antin, Joshua Kravitz, and Emil Martayan. "Detecting pneumonia in chest X-Rays with supervised learning." *Semanticscholar.org*, 2017.

[AM20] Ioannis D Apostolopoulos and Tzani A Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." *Physical and Engineering Sciences in Medicine*, p. 1, 2020.

[AYH20] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases." *Radiology*, p. 200642, 2020.

[CMD20] Joseph Paul Cohen, Paul Morrison, and Lan Dao. "COVID-19 image data collection." *arXiv preprint arXiv:2003.11597*, 2020.

[CRK20] Muhammad EH Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandakar R Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al-Emadi, et al. "Can AI help in screening viral and COVID-19 pneumonia?" *arXiv preprint arXiv:2003.13145*, 2020.

[DJV14] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. "Decaf: A deep convolutional activation feature for generic visual recognition." In *International conference on machine learning*, pp. 647–655, 2014.

[HZR16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[KGC18] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." *Cell*, **172**(5):1122–1131, 2018.

[KSH12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[KZG18] Daniel Kermany, Kang Zhang, and Michael Goldbaum. "Labeled optical coherence tomography (oct) and chest X-ray images for classification." *Mendeley data*, **2**, 2018.

[LBB98] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE*, **86**(11):2278–2324, 1998.

[NKP20] Ali Narin, Ceren Kaya, and Ziynet Pamuk. "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks." *arXiv preprint arXiv:2003.10849*, 2020.

[Org20] World Health Organization. "Coronavirus disease (COVID-19) outbreak situation." `https://www.who.int/emergencies/diseases/novel-coronavirus-2019`, 2020. [Online; accessed 28-May-2020].

[PY09] Sinno Jialin Pan and Qiang Yang. "A survey on transfer learning." *IEEE Transactions on knowledge and data engineering*, **22**(10):1345–1359, 2009.

[Rad] Radiopaedia Blog Rss. "Cases." `https://radiopaedia.org/cases`. [Online; accessed 27-May-2020].

[RDS15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision (IJCV)*, **115**(3):211–252, 2015.

[SAB20] Sana Salehi, Aidin Abedi, Sudheer Balakrishnan, and Ali Gholamrezanezhad. "Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients." *American Journal of Roentgenology*, pp. 1–7, 2020.

[SAS14] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. "CNN features off-the-shelf: an astounding baseline for recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 806–813, 2014.

[SCD17] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[SLJ15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[SZ14] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556*, 2014.

[UKK20] Buddhisha Udugama, Pranav Kadhiresan, Hannah N Kozlowski, Ayden Malekjahani, Matthew Osborne, Vanessa YC Li, Hongmin Chen, Samira Mubareka, Jonathan B Gubbay, and Warren CW Chan. "Diagnosing COVID-19: the disease and tools for detection." *ACS nano*, **14**(4):3822–3835, 2020.

[Wik20] Wikipedia contributors. "Coronavirus disease 2019 — Wikipedia, The Free Encyclopedia." `https://en.wikipedia.org/w/index.php?title=Coronavirus_disease_2019&oldid=959365303`, 2020. [Online; accessed 28-May-2020].

[WW20] Linda Wang and Alexander Wong. "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images." *arXiv preprint arXiv:2003.09871*, 2020.

[YYS] Y Yang, M Yang, C Shen, F Wang, J Yuan, J Li, M Zhang, Z Wang, L Xing, J Wei, et al. "Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. MedRxiv 2020: 2020.02. 11.20021493." *Sólo en caso de urgencia y si no se cuenta con otro donante podrá ser recolectado evaluando cuidadosamente los riesgos y beneficios. Podría volver a ser elegido si no tuvo historia de infección respiratoria severa, hayan trascurrido*, **28**.

[ZF14] Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks." In *European conference on computer vision*, pp. 818–833. Springer, 2014.